SUPPLEMENTARY NOTE

Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes

Kevin W. Kelley[1-5], Hiromi Nakao-Inoue[2-4], Anna V. Molofsky[2-4], Michael C. Oldham[1-3*]

[1]Department of Neurological Surgery, University of California at San Francisco, San Francisco, CA, USA
[2]The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California at San Francisco, San Francisco, CA, USA
[3]Weill Institute for Neurosciences, University of California at San Francisco, San Francisco, CA, USA
[4]Department of Psychiatry, University of California at San Francisco, San Francisco, CA, USA
[5]Medical Scientist Training Program and Neuroscience Graduate Program, University of California at San Francisco, San Francisco, CA, USA

*Correspondence: Michael.Oldham@ucsf.edu

Table of Contents

1. Supplementary discussion

1.1 Reproducibility of gene coexpression modules enriched with markers of major cell classes in CNS transcriptomes derived from intact tissue samples

We previously discovered highly reproducible gene coexpression modules in microarray data from intact human brain samples that were significantly enriched with markers of major CNS cell classes[1]. These findings were subsequently replicated by independent gene coexpression studies of intact CNS transcriptomes from mice[2-5], rats[6, 7], zebra finches[8, 9], macaques[10, 11], and humans[4, 12-17]; see Oldham[18] for further discussion. The reproducibility of these modules is the inevitable result of two simple ideas: i) different cell classes express different genes, and ii) intact tissue samples exhibit variability in cellular composition. Therefore, the genes that are most specifically and consistently expressed in the same cell class appear highly correlated when analyzed over a large number of biological replicate samples. Conversely, the expression patterns of these (high-fidelity) genes can be used to infer variation in the relative abundance of a cell class over heterogeneous samples, which can in turn be used to create mathematical models of gene expression as a function of variation in cellular composition. Below we provide additional discussion for diverse applications of this approach that we explore in the main text.

1.2 Identifying cellular and molecular correlates of aging

Several studies have reported age-related changes in gene expression in bulk human brain transcriptomes[19-21]. However, it has been difficult to determine whether such changes are primarily cell-intrinsic or due to changes in the cellular composition of the human brain with age. We therefore used expression patterns of high-fidelity genes to infer relationships between estimated cellular abundance and age in 32 CNS datasets. We observed that neuronal and oligodendroglial abundance were negatively correlated with age, while astrocytic and microglial abundance were positively correlated with age. The relative significance of the observed trends suggests that loss of neurons and oligodendrocytes may explain the relative increase in microglia and astrocytes. This interpretation is consistent with stereological evidence for decreased neuronal and oligodendroglial abundance in aged human brains[22, 23]. After controlling for variation in the inferred abundance of major cell classes, incorporating age as a covariate did not noticeably improve model performance (**Fig. 5E**). Furthermore, we were unable to identify genes that consistently showed age-related changes in gene expression that were independent of changes in cellular composition in all analyzed datasets (data not shown). While it is possible that our studies were underpowered to identify such changes, the most parsimonious explanation for our findings is that age-related changes in gene expression in bulk human brain samples are primarily driven by age-related changes in cellular composition.

1.3 Inferring cell-class associations for CNS disease genes

By modeling gene expression in bulk CNS transcriptomes as a function of variation in cellular composition, we inferred cell-class associations for diverse CNS disease genes. Some well-studied genes showed surprising associations. For example, mutations in *NPC1* cause Niemann-Pick type C disease, a rare autosomal recessive neurodegenerative disorder[24]. Studies of *NPC1* in model organisms have traditionally assumed a neuronal origin for disease pathogenesis; however, our analyses suggest that the overwhelming majority of *NPC1* mRNA in the adult human CNS is produced by oligodendrocytes (**Fig. 3B**). Using a curated database of results from genetic association studies[25], we analyzed cell-class-specific expression patterns of genes associated with complex CNS diseases. In general, genes expressed by neurons and astrocytes were associated with neurodevelopmental disorders, whereas genes expressed by astrocytes and microglia were associated with neurodegenerative disorders. The prevalence of significant

associations between astrocytic gene expression and genetic risk for developing diverse CNS diseases was unexpected and points to the need for more research on astrocyte biology in health and disease.

1.4 Determining molecular and cellular phenotypes in pathological samples

Gene expression modeling in bulk tissue transcriptomes can also reveal cell-class-specific expression changes associated with disease[26]. Using expression patterns of high-fidelity genes as proxies for cellular abundance, we controlled for variation in cellular composition and identified specific genes and biological pathways that were consistently up-regulated in AD neurons and microglia in independent datasets. For example, despite overall loss of neurons in AD, *HIGD1A* and *YWHAH* are significantly up-regulated in AD neurons (**Fig. 6G**). Interestingly, both of these genes have been shown to exert anti-apoptotic effects in response to cell stress[27, 28]. Genes that are up-regulated in AD microglia are also of particular interest given the strong association between AD risk alleles and microglial gene expression illustrated in **Fig. 6B**. Beyond identifying cell-intrinsic changes in gene expression, our approach also enables inferences about changes in cellular composition associated with pathology. We found that predicted changes in AD (decreased neuronal abundance and increased astrocytic / microglial abundance) were highly consistent across datasets. Although these changes are relative and must be confirmed by independent means, this strategy should accelerate efforts to determine whether differences in cellular composition are associated with diverse CNS disorders.

1.5 Studying regional diversity of human CNS cell classes

Although regional heterogeneity among neurons is well established, the extent of regional heterogeneity among glia is less well understood. By modeling gene expression in distinct CNS regions as a function of variation in cellular composition, we confirmed extensive transcriptional heterogeneity in neurons and identified nearly 100 genes whose expression patterns are likely to distinguish neuronal subtypes in the human CNS. We also found evidence for significant regional expression variation in astrocytes, which, to our knowledge, has not previously been described in the human brain. In contrast to previous work[29, 30], we observed much less evidence for expression variation among microglia and oligodendrocytes. However, it is important to note that our analysis was designed to detect binary expression differences. Analysis of more subtle differences in gene expression levels may therefore reveal additional evidence of regional diversity.

1.6 Identifying species differences in transcriptional regulation

Previous studies have compared CNS gene expression between humans and mice using *in situ* hybridization[31], gene coexpression analysis[4, 13, 32], or expression profiling of purified cell classes[33]. Our study extends previous work by employing a modeling framework and analyzing a larger number of samples, CNS regions, and technology platforms, while incorporating outgroup data from non-human primates. Over all homologous genes, neuronal expression fidelity was more conserved between humans and mice than glial expression fidelity. On a gene-by-gene basis, the majority of cell-class-specific expression differences were found in glia. Our findings are consistent with previous studies comparing humans and mice that reported weaker conservation of glial coexpression modules than neuronal coexpression modules[4, 13]. In contrast, analysis of purified human and mouse brain cell classes suggested similar transcriptional divergence for neurons and glia[33]. Although this discrepancy requires further study, we note that the strong conservation of expression fidelity in neurons relative to glia is mirrored at the protein level: high-fidelity neuronal genes are significantly less tolerant to loss-of-function and missense mutations than high-fidelity glial genes (**Fig. 3A-D**). Furthermore, resampling indicated that, on average, LoF (and missense) mutation intolerance for high-fidelity genes was far greater than expected by chance

for neurons ($p < 10\text{-}5$), but not glia ($p > 0.05$). Collectively, these findings suggest that neurons are under greater evolutionary constraint than glia.

1.7 New functional insights into human neurobiology

Human astrocytes are larger and more morphologically complex than mouse and non-human primate astrocytes[34, 35]. Upon transplantation of human glial progenitors into mice, one group found that human cells grew into large astrocytes and that the transplanted mice had enhanced cognitive abilities[36]. However, little is known about the cell-intrinsic molecular programs that govern these observed differences. Through comparative modeling of gene expression in humans and mice, we identified *PMP2* as highly expressed in human but not mouse astrocytes (**Fig. 8B-D,F**). We also found that expression of *PMP2* in the neocortex has increased monotonically from mice to non-human primates to humans (**Fig. 8E**), and that ectopic expression of *PMP2* in mouse astrocytes led to larger and more complex mouse astrocytes *in vivo* (**Fig. 8G-I, Fig. S12F**).

PMP2 is a small, basic, fatty acid-binding protein (a.k.a. *FABP8*) that localizes to membranes, binds lipids in its beta-barrel pocket, and attaches to negatively charged membranes through electrostatic and hydrophobic interactions[37, 38]. It is thought to be an important determinant of membrane stability and lipid dynamics in peripheral myelin[39] and has been shown to transport fatty acids to lipid vesicles and membranes[39]. Molecular simulations have identified cholesterol as one major candidate for binding in the PMP2 pocket[37]. Because cholesterol and other fatty acids are important for membrane fluidity, we postulate that *PMP2* expression has increased the morphological complexity of human astrocytes by enhancing membrane fluidity, thereby expanding their volumetric domain and the plasticity with which they are able to monitor and respond to their extracellular environments.

REFERENCES

1. Oldham, M.C., *et al.* Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271-1282 (2008).

2. Fertuzinhos, S., *et al.* Laminar and temporal expression dynamics of coding and noncoding RNAs in the mouse neocortex. *Cell Rep* **6**, 938-950 (2014).

3. Iancu, O.D., *et al.* Genetic diversity and striatal gene networks: focus on the heterogeneous stock-collaborative cross (HS-CC) mouse. *BMC Genomics* **11**, 585 (2010).

4. Miller, J.A., Horvath, S. & Geschwind, D.H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *PNAS* **107**, 12698-12703 (2010).

5. Mulligan, M.K., *et al.* Molecular profiles of drinking alcohol to intoxication in C57BL/6J mice. *Alcohol. Clin. Exp. Res.* **35**, 659-670 (2011).

6. Ponomarev, I., Rau, V., Eger, E.I., Harris, R.A. & Fanselow, M.S. Amygdala transcriptome and cellular mechanisms underlying stress-enhanced fear learning in a rat model of posttraumatic stress disorder. *Neuropsychopharmacology* **35**, 1402-1411 (2010).

7. Winden, K.D., *et al.* A systems level, functional genomics analysis of chronic epilepsy. *PLoS One* **6**, e20763 (2011).

8. Hilliard, A.T., Miller, J.E., Fraley, E.R., Horvath, S. & White, S.A. Molecular microcircuitry underlies functional specification in a basal ganglia circuit dedicated to vocal learning. *Neuron* **73**, 537-552 (2012).

9. Hilliard, A.T., Miller, J.E., Horvath, S. & White, S.A. Distinct neurogenomic states in basal ganglia subregions relate differently to singing behavior in songbirds. *PLoS Comput. Biol.* **8**, e1002773 (2012).

10. Bernard, A., *et al.* Transcriptional architecture of the primate neocortex. *Neuron* **73**, 1083-1099 (2012).

11. Bakken, T.E., *et al.* A comprehensive transcriptional map of primate brain development. *Nature* **535**, 367-375 (2016).

12. Ben-David, E. & Shifman, S. Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. *PLoS genetics* **8**, e1002556 (2012).

13. Hawrylycz, M., *et al.* Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832-1844 (2015).

14. Podtelezhnikov, A.A., *et al.* Molecular insights into the pathogenesis of Alzheimer's disease and its relationship to normal aging. *PLoS One* **6**, e29610 (2011).

15. Ponomarev, I., Wang, S., Zhang, L., Harris, R.A. & Mayfield, R.D. Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. *J. Neurosci.* **32**, 1884-1897 (2012).

16. Torkamani, A., Dean, B., Schork, N.J. & Thomas, E.A. Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res.* **20**, 403-412 (2010).

17. Voineagu, I., *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380-384 (2011).

18. Oldham, M.C. Transcriptomics: from differential expression to coexpression. in *The OMICs: applications in neurosciences.* (ed. G. Coppola) 85-113 (Oxford, 2014).

19. Lu, T., *et al.* Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883-891 (2004).

20. Kumar, A., *et al.* Age-associated changes in gene expression in human brain and isolated neurons. *Neurobiol. Aging* **34**, 1199-1209 (2013).

21. Lu, T., *et al.* REST and stress resistance in ageing and Alzheimer's disease. *Nature* **507**, 448-454 (2014).

22. Pelvig, D.P., Pakkenberg, H., Stark, A.K. & Pakkenberg, B. Neocortical glial cell numbers in human brains. *Neurobiol. Aging* **29**, 1754-1762 (2008).

23. Soreq, L., *et al.* Major Shifts in Glial Regional Identity Are a Transcriptional Hallmark of Human Brain Aging. *Cell Rep* **18**, 557-570 (2017).

24. Carstea, E.D., *et al.* Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* **277**, 228-231 (1997).

25. Yu, W., Clyne, M., Khoury, M.J. & Gwinn, M. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics* **26**, 145-146 (2010).

26. Kuhn, A., Thu, D., Waldvogel, H.J., Faull, R.L. & Luthi-Carter, R. Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. *Nat Methods* **8**, 945-947 (2011).

27. An, H.J., *et al.* The survival effect of mitochondrial Higd-1a is associated with suppression of cytochrome C release and prevention of caspase activation. *Biochim. Biophys. Acta* **1813**, 2088-2098 (2011).

28. Buret, L., *et al.* Loss of function of Ywhah in mice induces deafness and cochlear outer hair cells' degeneration. *Cell Death Discov* **2**, 16017 (2016).

29. Grabert, K., *et al.* Microglial brain region-dependent diversity and selective regional sensitivities to aging. *Nat. Neurosci.* **19**, 504-516 (2016).

30. Marques, S., *et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326-1329 (2016).

31. Zeng, H., *et al.* Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149**, 483-496 (2012).

32. Lui, J.H., *et al.* Radial glia require PDGFD-PDGFRbeta signalling in human but not mouse neocortex. *Nature* **515**, 264-268 (2014).

33. Zhang, Y., *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37-53 (2016).

34. Oberheim, N.A., Goldman, S.A. & Nedergaard, M. Heterogeneity of astrocytic form and function. *Methods Mol. Biol.* **814**, 23-45 (2012).

35. Oberheim, N.A., *et al.* Uniquely hominid features of adult human astrocytes. *J. Neurosci.* **29**, 3276-3287 (2009).

36. Han, X., *et al.* Forebrain engraftment by human glial progenitor cells enhances synaptic plasticity and learning in adult mice. *Cell Stem Cell* **12**, 342-353 (2013).

37. Majava, V., *et al.* Structural and functional characterization of human peripheral nervous system myelin protein P2. *PLoS One* **5**, e10300 (2010).

38. Ruskamo, S., *et al.* Atomic resolution view into the structure-function relationships of the human myelin peripheral membrane protein P2. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 165-176 (2014).

39. Zenker, J., *et al.* A role of peripheral myelin protein 2 in lipid homeostasis of myelinating Schwann cells. *Glia* **62**, 1502-1512 (2014).